Jñānābha, Vol. 49(2) (2019), 6-14

AIR QUALITY PREDICTION USING TIME SPACE ANALYSIS

By

Rashmi Bhardwaj* **and Dimple Pruthi**

Non Linear Dynamics Lab, University School of Basic and Applied Sciences

Guru Gobind Singh Indraprastha University, Delhi-110078, India

Email:*rashmib22@gmail.com

(*Received : August* 02, 2019 *; Revised: October* 18, 2019)

Abstract

Air pollution is a serious threat to the environment and ecology. Monitoring and prediction of air quality is an important aspect, as it helps to issue early warnings and adopt suitable control measures in time. Particulate matter of size less than and equal to 2.5 microns is the prominent air pollutant. It easily penetrates through lungs affecting human health. This paper investigates the performance of the empirical mode decomposition and the wavelet transform in non linear non stationary PM2.5 time series prediction problem. The prediction is carried out by applying adaptive neuro-fuzzy inference system (ANFIS). It is found that the wavelet transform outperforms empirical mode decomposition for non linear PM2.5 time series.

2010 Mathematics Subject Classification: 62P12

Keywords and phrases: Particulate matter, Air quality index, Empirical mode decomposition, Wavelet transform, Adaptive neuro-fuzzy inference system.

1 Introduction

Air pollutants are growing rapidly, which is one of the most serious health issues. Among pollutants, particulate matter is gaining more attention because of its ability to penetrate deep into lungs. Many studies have been carried out showing significant association between particulate matter and chronic diseases and even resulting in death [2,7]. The WHO has called it second tobacco epidemic. The PM2.5 can hamper brain development by 10-20 per cent and we are going to make our next generation retarded. Evidence is now emerging that heart attacks, a brain attacks are linked to air pollution. A recent study found strong relation between PM2.5 exposure and neurological disorders [9]. It has aroused worldwide concerns over the last few years. The study of particulate matter is of great relevance from worlds economy and health point of view. The spatial and temporal distribution of PM2.5 concentration which has strong nonlinear characteristics is influenced by the meteorological field, the emission source, the complex underlying surface, and the coupling of the physical and chemical processes, which makes it difficult to predict. At present, the commonly used prediction methods are mechanism analysis and statistical prediction method. Many studies related to prediction of PM2.5 has been done using stepwise regression keeping in mind its chemical composition and physical properties, and the other pollutants (such as SO_2 ,NO_x, O_3) and meteorological factors [3,4,8]. PM10 and PM2.5 in urban areas using a chemical transport model is predicted [11]. In this paper, PM2.5 prediction is carried out by breaking down time series into sub series. The study examines Wavelet-ANFIS (WANFIS) and Empirical mode decomposition-ANFIS (EMD-ANFIS) models for predicting PM2.5 using the embedded series.

The proposed model is trained for PM2.5 concentration monitored at Shadipur, Delhi. Shadipur is one of the pollution hotspots in Delhi. It is residential cum industrial area. Delhi Transport Corporation depot is also located in Shadipur connecting many parts of Delhi. Its neighboring areas are Mayapuri and Naraina. Naraina is divided into industrial, residential and rural areas. It is the location of the headquarters of the Steel Authority of India Limited. Mayapuri is combination of residential flats, light metal factories and automobile service stations. Air quality in this area lies mainly in very poor or severe category. Apart from meteorological parameters and vehicular emissions, industries plays major role in pollution levels making study of particulate matter more complex. The main cause of air pollution is garbage burning. The garbage predominantly consists of rubber and plastic waste which is carcinogenic. PM2.5 (24-hourly) concentrations are obtained from Central Pollution Control Board (CPCB) from the period from March, 2015 to March, 2019.

2 Time space analysis method

The aim is to forecast PM2.5 daily concentration using past values. The uncertainty of PM2.5 makes precise prediction a challenging task. Alternate way is to break the original series into components with lower variability and then training each of them using ANFIS. Summing individual results gives the final output [12,14]. The time space analysis methods-wavelet transformation and empirical mode decomposition are used for breaking down high variability time series into subseries. Further, performance analysis of both methods is carried out by comparing PM2.5 actual and forecasted values.

2.1 Empirical mode Decomposition (EMD)

The EMD decomposes a signal into intrinsic membership functions (IMF). IMF has only one extreme between zero crossings, and has an average of zero [6]. The process used for decomposing is iterative and is stopped when the standard deviation between two successive shifting is smaller than 0.2 or 0.3. Given a signal $y(t)$ shifting process works in following steps:

1. From cube-spline interpolation of local extremes determine upper and lower envelope respectively. Let μ_1 be the average of upper and lower envelope.
2. First component, $A_1(t) = y(t) - \mu_1(t)$.
3. If μ_1 and A_1 satisfy stopping criteria, then first IMF $d_1(t) = \mu_1(t)$ and residue $R_1(t) = A_1(t)$.
4. Otherwise, steps 1-3 is repeated for $A_1(t)$.
5. For $R_1(t)$, steps 1-4 are repeated until all IMFs and residue are obtained say J is IMF count

$$y(t) = \Sigma_{i=1}^{J} D_i(t) + R(t). \tag{2.1}$$

2.2 Wavelet Transform

Fourier transforms limitation of uniform frequency resolution at all frequencies lead to wavelet transforms. With Fourier transforms time-frequency grid is uniform whereas wavelet transform is collection of different windows thus descrying low and high frequencies. Wavelet transforms are further divided into continuous and discrete [1]. The present study deals with discrete wavelet transform (DWT) for multi resolution of signal. Multi resolution analysis (finer to coarser in time domain) is based on high and low pass filter. High pass filter referred

to as mother wavelet captures high frequency (details) and low pass filter are father wavelet which captures low frequency (approximations) [10]

$$\phi(t) = \sqrt(2) \sum g_a \phi(2t - a), g_a = \frac{1}{\sqrt(2)} \int \phi(t)\phi(2t - a)dt, \tag{2.2}$$

$$\psi(t) = \sqrt(2) \sum f_a \psi(2t - a), f_a = \frac{1}{\sqrt(2)} \int \psi(t)\psi(2t - a)dt. \tag{2.3}$$

Using mother and father wavelet, signal x(t) can be disintegrated as:

$$x(t) = \Sigma_c X_{l,c}\psi_{l,c}(t) + \Sigma_c d_{l,c}\psi_{l,c}(t) + \Sigma_c d_{l-1,c}\psi_{l-1,c}(t) + ... + \Sigma_c d_{1,k}\psi_{1,c}, \tag{2.4}$$

$$X_{l,c} = \int \phi_{l,c}x(t)dt, \phi_{l,c} = 2^{\frac{-l}{2}} \phi(2^{-l}t - c), \tag{2.5}$$

$$d_{l,c} = \int \psi_{l,c}x(t)dt, \psi_{l,c} = 2^{\frac{-l}{2}} \psi(2^{-l}t - c). \tag{2.6}$$

Here, l ranges from 1 to number of number of coefficients and c is number of levels. High and low pass filter vary for different wavelets. For present study, daubechies wavelet due to its decay pattern in time and frequency unlike Haar and Shannon wavelets which are compact in time and in frequency respectively.

3 ANFIS

ANFIS is based on five adaptive and fixed layers comprising of premise and consequent parameters. The premise parameters are the input parameters defined by the membership functions of data. The consequent parameters are tuned. The learning ability of neural networks helps in tuning the parameters. ANFIS is basically used for handling non linear behavior of time series. The algorithm of the process learns in accordance with the back propagation method. Many prediction studies based on data is carried out using ANFIS [5]. The performance of the system in learning from data helps in predicting non stationary non linear time series. More insights to the system architecture are discussed [13].

4 Methodology

PM2.5 concentration monitored at Shadipur is undertaken for the study for the period from 22/03/2015 to 28/03/2019. The original series is decomposed using time space analysis methods. Consider time series of PM2.5 concentration as $x(t)$, $t = 1, 2, ..., m$ where m is the count of data points under consideration. The optimal lag undertaken in for the prediction is four keeping in mind the learning rate which will directly affect the computational time. The first step is to decompose series than apply ANFIS on decomposed components for prediction as depicted in Fig. 5.1.

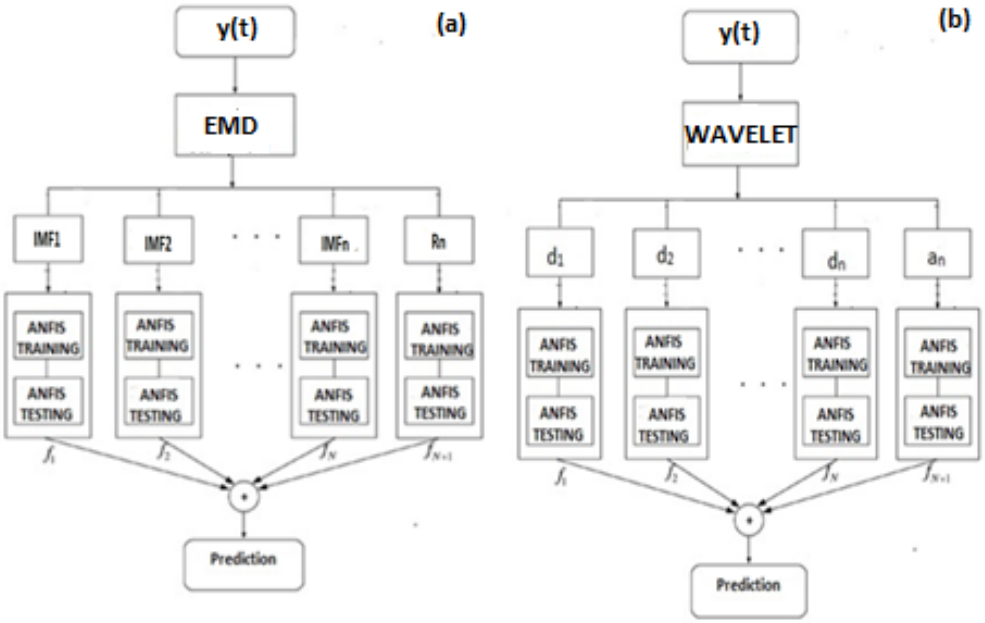

Figure 4.1: (a) EMD-ANFIS (b) Wavelet ANFIS (WANFIS)

5 Performance Testing

The section comprises of performance analysis of proposed EMD-ANFIS and WANFIS algorithm is tested for PM2.5 series. The data is split into training (70 percent), testing (15 percent) and validation (15 percent) sets. The performance has been compared between the observed and predicted value using mean absolute error (MAE) and coefficient of determination (R^2). R^2 close to 1 and smaller value of MAE proves good prediction of the methods. where, a_i is actual data, pi predicted data, m data points and a is mean of the actual data. Further to investigate the validated data, air quality index (AQI) is evaluated and compared with the actual as depicted on CPCB site on daily basis. AQI is not just a number describing the quality of air but also explains what we are breathing in. The purpose was to acquaintance public about decaying air quality. AQI is calculated using the standards defined by CPCB. AQI is broadly categorized as follows:

Category	Value	24-hr Average $PM_{2.5}$ Concentration
Good	0-50	0-30
Satisfactory	51-100	31-60
Moderately polluted	101-150	61-90
Poor	151-200	91-120
Very poor	201-300	121-250
Severe	301-500	250+

The time series of PM2.5 for the period under consideration is depicted in Figure 5.1 and plot area describes PM2.5 categories. The series clearly depicts few cases of less PM2.5

concentration and more of poor and very poor categories. Firstly the series is decomposed using EMD. IMF is shown in Figure 5.2.

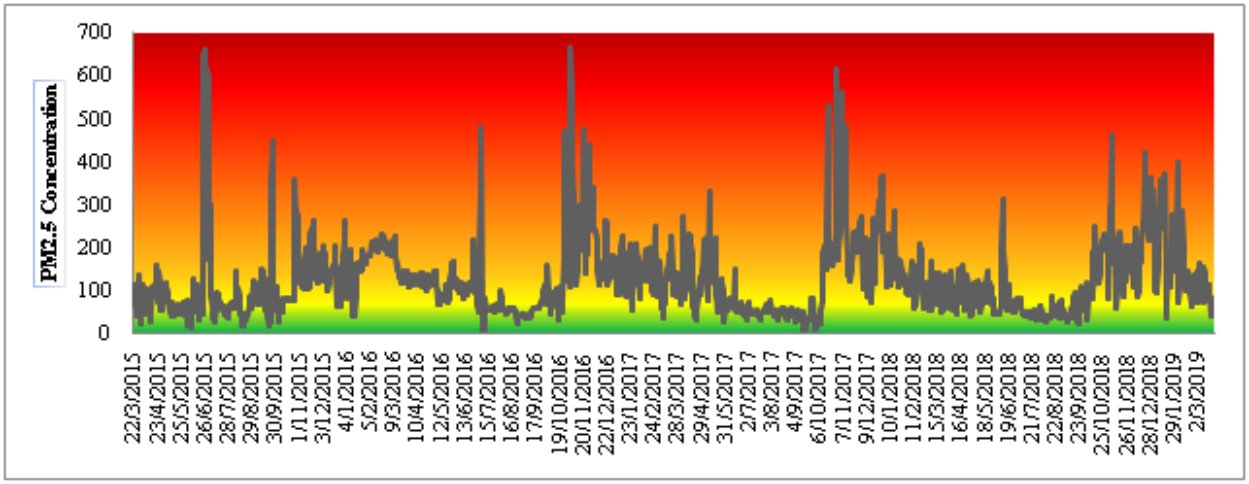

Figure 5.1: PM2.5 time series for the period from 22/03/2015 to 28/03/2019.

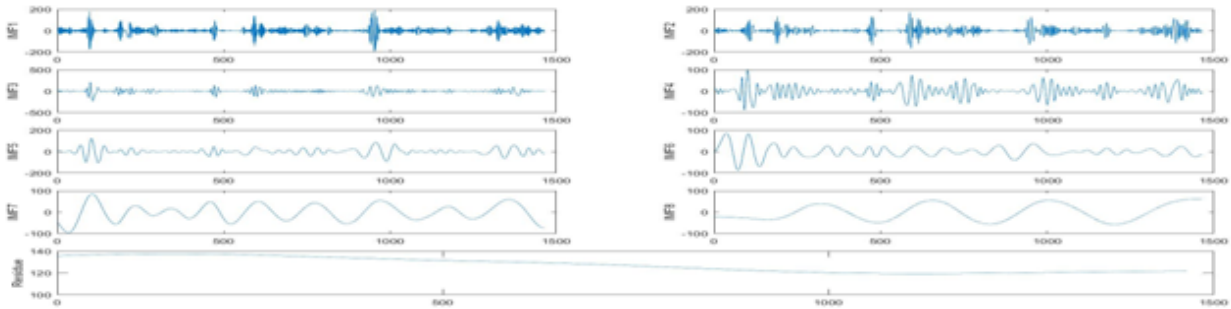

Figure 5.2: IMFs/ Residue series of PM2.5 data.

The decomposed IMFs and residue are predicted using ANFIS. Next PM2.5 series undergo wavelet transformation and is decomposed into approximations and details which follow the algorithm depicted in Figure 4.1b. Finally the predicted data is obtained by summing up predicted value of components. The performance of models can be seen in Figure 5.3 and Figure 5.4 for observed and predicted value of testing and validation data.

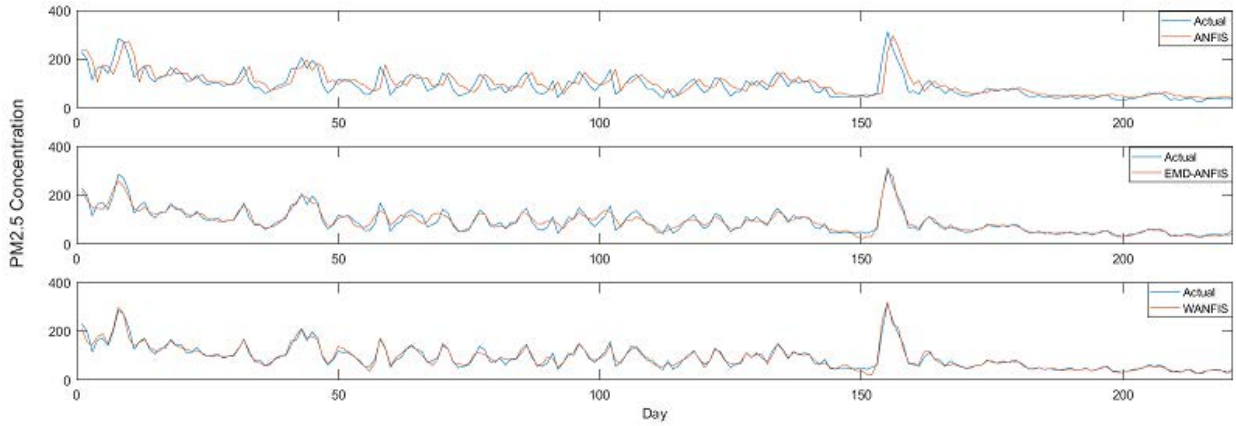

Figure 5.3: Observed and predicted PM2.5 Testing data.

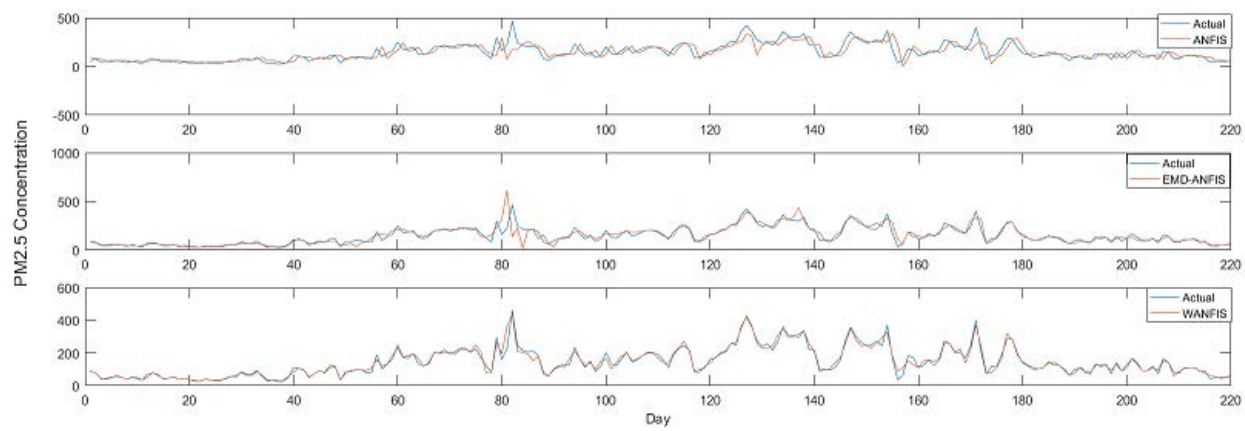

Figure 5.4: Observed and predicted PM2.5 Validation data.

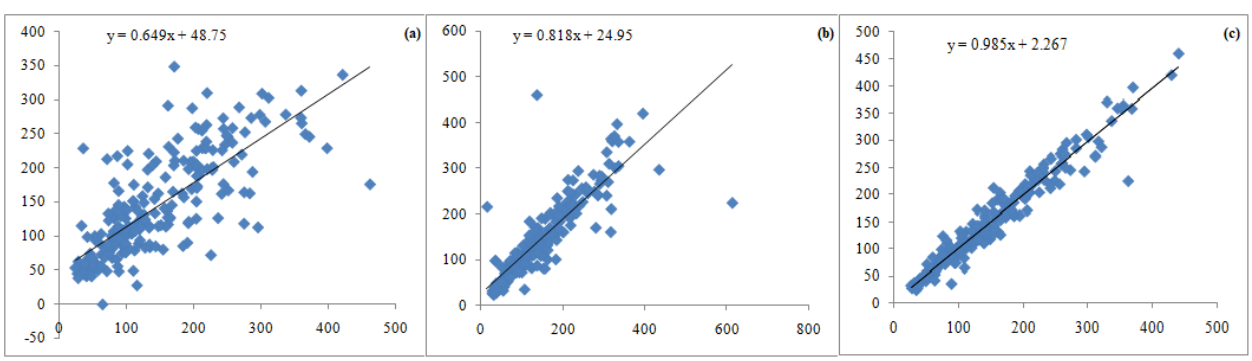

Figure 5.5: Observed vs. predicted PM2.5 using(a)ANFIS;(b)EMDANFIS;(c)WaveletANFIS

Further, the efficiency of models is statistically compared using MAE and R^2 as depicted in Table 5.1.

Table 5.1: Performance Analysis

Model	R^2			MAE		
	Training	Testing	Validation	Training	Testing	Validation
ANFIS	0.7324	0.6251	0.5814	31.19	27.9552	33.125
EMD ANFIS	0.9491	0.9079	0.7163	14.4091	11.6029	25.7133
Wavelet ANFIS	**0.9762**	**0.9507**	**0.9505**	**9.5249**	**8.0686**	**13.0044**

It is evident that WANFIS performs much better than other models. Using the forecasted and observed validation data AQI is calculated. Approximately 96 percent of forecasted and observed AQIs lie in the same category as in Figure 5.6. Thus the model can be used by pollution controlling and monitoring agencies to predict AQI which can contribute in controlling its severe affects and taking precaution beforehand by the government and people.

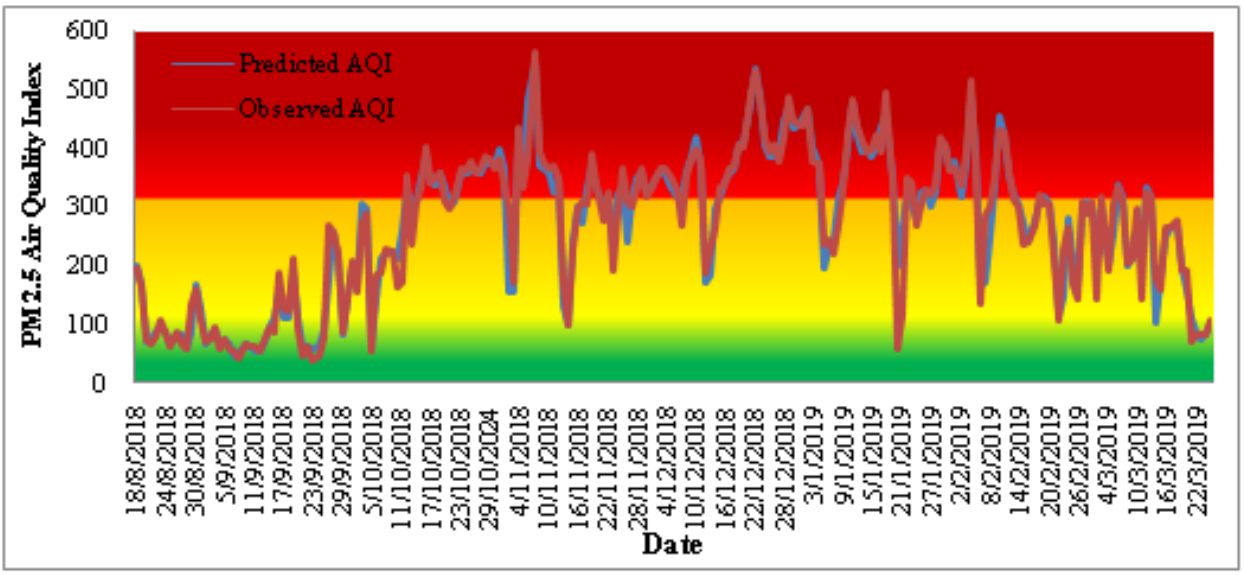

Figure 5.6: Predicted and observed air quality index, plot area is divided into categories.

6 Conclusion

A hybrid model consisting of time space analysis methods and adaptive neuro-fuzzy inference system is proposed for accurate prediction. The performance of the model has been verified by calculating AQI of the predicted values and comparing with AQIs for PM2.5 as shown on www.cpcb.nic.in. It is found that the wavelet transformation is competent to improve the PM2.5 forecasting accuracy. Simulation results have shown that Wavelet-ANFIS outperformed EMD-ANFIS and ANFIS. Air quality index falls in poor and severe categories indicating need of action plan to control air pollution. Wavelet-ANFIS lead to accurate prediction of AQI. WANFIS can be used in order to forecast air quality index and issuing health advisories according to the air quality categories. The proposed model can be considered for prediction of any non linear non stationary time series using only the lagged values of the concerned series. The model reduced dependency complexity on other factors. Although the work reported in this paper improves prediction accuracy, there is still ample space for improvement in learning rate of the proposed model. For future studies authors

will consider other optimization techniques to optimize learning rate and computational time for more convenient model in forecasting air quality.

Acknowlegements
The authors are thankful to Guru Gobind Singh Indraprastha University, Delhi (India) for providing research facilities and financial support. They are also thankful to the referee for his valuable suggestions.

References

[1] R. Bhardwaj, Wavelets and fractal methods with environmental applications, A. H. Siddiqi, P. Manchanda and R. Bhardwaj (Eds), *Mathematical Models, Methods and Applications*, 173-195, Springer, 2016.

[2] H. Chen, M. S. Goldberg and P.J. Villeneuve, A systematic review of the relation between long-term exposure to ambient air pollution and chronic diseases, *Reviews on Environmental Health*, **23(4)** (2008), 243-97.

[3] X. Feng, Q. Li, Y. Zhu, J. Hou, L. Jin and J. Wang, Artificial neural networks forecasting of PM2.5 pollution using air mass trajectory based geographic model and wavelet transformation, *Atmospheric Environment*, **107**(2015), 118-128.

[4] H-d. He, M. Li, W. Wang, Z. Wang and Y. Xue, Prediction of PM2.5 concentration based on the similarity in air quality monitoring network, *Building and Environment*, **137** (2018), 11-17.

[5] J. S. R. Jang, ANFIS Adaptive-Network-based Fuzzy Inference System, *IEEE Transactions on Systems Man and Cybernetics*, **23(3)** (1993), 665-685.

[6] L. Karthikeyan and D. N. Kumar, Predictability of non stationary time series using wavelet and EMD based ARMA models, *Journal of Hydrology*, **502** (2013), 103-119.

[7] T. Li, R. Hu, Z. Chen, Q. Li, S. Huang, Z. Zhu and L. F. Zhou, Fine particulate matter (PM2.5): The culprit for chronic lung diseases in China, *Chronic Diseases and Translational Medicine*, **4(3)** (2018), 176-186.

[8] J. B. Ordieres, E. P. Vergara, R. S. Capuz and R. E. Salazar, Neural network prediction model for fine particulate matter (PM2.5) on the US-Mexico border in El Paso (Texas) and Ciudad Juarez (Chihuahua), *Environmental Modelling and Software*, **20** (2005), 547-559.

[9] F. Pengfei, G. Xinbiao, C. Felix Man Ho and Y. Ken Kin Lam, The association between PM2.5 exposure and neurological disorders: A systematic review and meta-analysis,*Science of the Total Environment*, **655** (2019), 1240-1248.

[10] T. Rajaee, S. A. Mirbagheri, V. Nourani and A. Alikhani, Prediction of daily suspended sediment load using wavelet and neurofuzzy combined model, *International Journal of Environmental Science and Technology*, **7(1)** (2010), 93-110.

[11] P. E. Saide, G. R. Carmichael, S. N. Spak, L. Gallardo, A. E. Osses, M. A. Mena-Carrasco and M. Pagowski, Forecasting urban PM10 and PM2.5 pollution episodes in very stable nocturnal conditions and complex terrain using WRF-Chem CO tracer model, *Atmospheric Environment*, **45(16)** (2011), 2769-2780.

[12] M. Shafaei and O. Kisi, Lake Level Forecasting Using Wavelet-SVR, WaveletANFIS and Wavelet-ARMA Conjunction Models, *Water Resources Management*, **30** (2016),79-97.

[13] A. Solgi, A. Pourhaghi, R. Bahmani and H. Zarei, Improving SVR and ANFIS performance using wavelet transform and PCA algorithm for modeling and predicting Biochemical Oxygen Demand (BOD),*Ecohydrology and Hydrobiology*, **134** (2016), 1-13.

[14] A. Yarar, A Hybrid Wavelet and Neuro-Fuzzy Model for Forecasting the Monthly Streamflow Data, *Water Resources Management*, **28** (2014), 553-565.